Homework #4

1. In this problem, you'll prove the shortened version of Eckhart-Young that we went over in class. Let X be an $n \times p$ matrix with svd decomposition $X = \sum_{i=1}^{\min(n,p)} s_i u_i v_i^T$. Let $B = \sum_{i=1}^{k} s_i u_i v_i^T$. and assume $s_k > s_{k+1}$ for a given k. We want to determine a solution to the minimization,
$$\min_{Y \sim n\times p, \mathrm{rank}(Y)=k} \|X - Y\|_2^2. \tag{1}$$
   (a) Show $\|X - B\|_2^2 = s_{k+1}^2$.

   (b) Suppose Y is rank k. Show that Y must have the form,
$$Y = \sum_{i=1}^{k} r_i a_i b_i^T, \tag{2}$$
   where $r_i \in \mathbb{R}$, $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}^p$.

   (c) Suppose
$$\mathrm{span}(b_1, b_2, \ldots, b_k) \neq \mathrm{span}(v_1, v_2, \ldots, v_k). \tag{3}$$
   Show that there exists $w \in \mathrm{span}(v_1, v_2, \ldots, v_k)$ with w orthogonal to every vector in $\mathrm{span}(b_1, b_2, \ldots, b_k)$. Then show $\|X - Y\|_2^2 \geq s_k^2$.

   (d) Using (c) show that if Y solves (1) then,
$$\mathrm{span}(b_1, b_2, \ldots, b_k) = \mathrm{span}(v_1, v_2, \ldots, v_k) \tag{4}$$
   and
$$\mathrm{span}(a_1, a_2, \ldots, a_k) = \mathrm{span}(u_1, u_2, \ldots, u_k). \tag{5}$$
   Then show that for this optimal Y, we must have $\|X - Y\|_2^2 \geq s_{k+1}^2$ and conclude that B is a solution to (1).

2. Consider again the MNIST dataset. Let X be the data matrix. Let x_i be the ith row of X, corresponding to the ith image. For all values of k, determine the fraction of the dataset's variance captured by a k-dimensional PCA. For a given k-dimensional PCA, let $x_i' \in \mathbb{R}^{784}$ be the PCA approximation (i.e. after decoding) of x_i. Consider image 1, which is a 5. Use different k to compute x_1' and visualize the projection using matplotlib's imshow function. For which k do you begin to see the 5? How does that relate to the fraction of variation captured by the PCA? (For this problem do not use sklearn's PCA. You may use numpy or scipy's eigenvalue or svd functions.)

3. This problem will set the scene for implementing ICA on the next homework. In this problem we will construct S and A matrices that we will then use to construct X. Recall $X \sim n \times p$, $A \sim n \times k$, and $S \sim k \times p$. In the next homework, we will pretend we only know X and see how well ICA reconstructs S and A.

   (a) Write a function sample_signal(p) that return a random vector $s \in \mathbb{R}^p$ as follows. s_1 should be chosen randomly from the values $\{1, 2, 3, 4\}$ with each value having equal probability. Then iteratively construct the s_i for $i = 2, 3, \ldots, p$ as follows. With probability 0.9, $s_i = s_{i-1}$ and with probability 0.1, s_i is chosen from $\{1, 2, 3, 4\}$ with each value having equal probability.

(b) Use sample_signal(p) with $p = 100$ to sample 5 signals and row bind them to form the S matrix which will be $5 \times p$. Plot the 5 signals, the rows of S, just to see their difference. Let $n = 100$. Decide on a sensible way to define the A matrix that linearly combines the S matrix to from the data matrix X. Add some noise to the data by defining,

$$X = AS + 0.1\eta, \tag{6}$$

where η is an $n \times p$ matrix with each entry given by a independent standard normals. (The point of adding η is that it will make X rank n. Without it X has the rank of S, which is 5, and in the steps below M will not be invertible. We'll discuss this more fully in class.)

(c) Write a function that performs the following two transforms on X and outputs the matrix Z:

   i. Let Y be the row centered version of X for which all rows have mean zero. (**Careful, in PCA we center the columns, but in ICA we center the rows**. I forgot to emphsize that in class.)

   ii. Let $\Sigma = \frac{1}{p} Y Y^T$. Note Σ is $n \times n$. Construct an invertible $n \times n$ matrix M such that $\Sigma = M M^T$. (M is often written as $\Sigma^{1/2}$.). Let $Z = M^{-1} Y$.

(d) Let $w \in \mathbb{R}^n$ with $\|w\| = 1$. Let $y = w^T Z$. Note that $y \in \mathbb{R}^p$. Show

   i. $\frac{1}{p} \sum_{i=1}^{p} y_i = 0$

   ii. $\frac{1}{p} \sum_{i=1}^{p} y_i^2 = 1$

(e) Let

$$G(x) = -\exp[-\frac{x^2}{2}] \tag{7}$$

If you've taken probability do part (i), if not do part (ii),

   i. Let X be a standard normal r.v. Compute $E[G(X)]$. Let $x_1, x_2, \dots, x_p$ be iid samples from X. Explain why $E[G(x)]$ is a good estimator of $\frac{1}{p} \sum_{i=1}^{p} G(x_i)$.

   ii. Determine,

$$\int_{-\infty}^{\infty} f(x) G(x) dx, \tag{8}$$

   where $f(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2}]$. Use numpy or scipy's normal sampler to sample 1 million standard normals (i.e. a normal with mean 0 and variance 1). Compare the value of the integral to $\frac{1}{p} \sum_{i=1}^{p} G(x_i)$ where x_i is the ith of your 1 million samples.